PCT/GB 99/02720
09/763019
REC'D 15 SEP 1999
WIPO PCT

The Patent Office

INVESTOR IN PEOPLE

# PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

GB99/2720

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed P. Mahoney

Dated 7 September 1999

Patents Form 1/77

Patents Act 1977
(Rule 16)

THE PATENT OFFICE
18 AUG 1998
RECEIVED

Patent Office

1/77

# Request for grant of a patent

*(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)*

18 AUG 1998

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

PPD 50355/GB/P

P01/7700 25.00 - 9818003.7

2. Patent application number
*(The Patent Office will fill in this part)*

9818003.7

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

ZENECA Limited
15 Stanhope Gate
London W1Y 6LN
UNITED KINGDOM

Patents ADP number *(if you know it)*

6254007002

If the applicant is a corporate body, give the country/state of its incorporation

UNITED KINGDOM

4. Title of the invention

POLYNUCLEOTIDE SEQUENCES

5. Name of your agent *(if you have one)*

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

Frank Mackie HUSKISSON
Intellectual Property Department
ZENECA Agrochemicals
Jealott's Hill Research Station
P O Box 3538
Bracknell Berkshire RG42 6YA
UNITED KINGDOM

Patents ADP number *(if you know it)*

03969060002

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

| Country | Priority application number *(if you know it)* | Date of filing *(day / month / year)* |
|---|---|---|

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

| Number of earlier application | Date of filing *(day / month / year)* |
|---|---|

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) *any applicant named in part 3 is not an inventor, or*

b) *there is an inventor who is not named as an applicant, or*

c) *any named applicant is a corporate body.*
*See note (d))*

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

| | |
|---|---|
| Continuation sheets of this form | |
| Description | 17 |
| Claim(s) | 4 |
| Abstract | — |
| Drawing(s) | 18 + 16 |

10. If you are also filing any of the following, state how many against each item.

| | |
|---|---|
| Priority documents | |
| Translations of priority documents | |
| Statement of inventorship and right to grant of a patent *(Patents Form 7/77)* | |
| Request for preliminary examination and search *(Patents Form 9/77)* | |
| Request for substantive examination *(Patents Form 10/77)* | |
| Any other documents *(please specify)* | |

11. I/We request the grant of a patent on the basis of this application.

ZENECA LIMITED

Signature VKMallipeddi Date 18TH AUGUST 1995

Authorised Signatory

12. Name and daytime telephone number of person to contact in the United Kingdom VIJAYA KUMARI MALLIPEDDI 01344 414365

**Warning**

*After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.*

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

# POLYNUCLEOTIDE SEQUENCES

The present invention relates to recombinant DNA technology, and in particular to nucleotide sequences (and expression products thereof) which are used in the production of transgenic plants.

The present invention provides, *inter alia*, nucleotide sequences useful in the production of plants which show improved resistance to infection by microorganisms such as bacteria and fungi.

According to the present invention there is provided a polynucleotide comprising a sequence selected from those depicted in SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 3, SEQ ID No. 4 and SEQ ID No. 5.

A particularly preferred embodiment of the polynucleotide consists of nucleotides 137 to 286 in SEQ ID No. 1, nucleotides 95 to 244 in SEQ ID No. 2, nucleotides 108 to 257 in SEQ ID No. 3, nucleotides 104 to 253 in SEQ ID No. 4 or nucleotides 177 to 326 in SEQ ID No. 5. Also included within the invention is the translation product of the region comprised by nucleotides 137 to 286 in SEQ ID No. 1, by nucleotides 95 to 244 in SEQ ID No. 2, by nucleotides 108 to 257 in SEQ ID No. 3 or by nucleotides 104 to 253 in SEQ ID No. 4 and protein having an amino acid sequence which is at least 95% similar to said product. The translation product is an antimicrobial protein. The related antimicrobial proteins DmAMP1 and Dm-AMP2 have been described in Published International Patent Application No. WO 93/05153.

A further preferred embodiment of the polynucleotide consists of nucleotides 287 to 385 in SEQ ID NO. 1, nucleotides 245 to 334 in SEQ ID No. 2, nucleotides 258 to 317 in SEQ ID No. 3, nucleotides 254 to 343 in SEQ ID No. 4 or nucleotides 327 to 446 in SEQ ID No.5. These nucleotides encode protein sequences which may be used as cleavable linkers in the co-expression of multiple proteins as is described further herein. The invention further extends to the translation product of nucleotides 287 to 385 in SEQ ID NO. 1, nucleotides 245 to 334 in SEQ ID No. 2, nucleotides 258 to 317 in SEQ ID No. 3, nucleotides 254 to 343 in SEQ ID No. 4 or nucleotides 327 to 446 in SEQ ID No.5 and protein having an amino acid sequence which is at least 85% similar to said product.

A yet further preferred embodiment of the polynucleotide consists of nucleotides 53 to 136 in SEQ ID No. 1, nucleotides 11 to 94 in SEQ ID No.2, nucleotides 24 to 107 in SEQ

ID No. 3, nucleotides 20 to 103 in SEQ ID No. 4 or nucleotides 1 to 176 in SEQ ID No. 5 excluding the sequence encoding the intron marked at positions 65 to 156. These nucleotide sequences are signal sequences which may be linked to homologous and heterologous protein encoding regions to transport proteins extracellularly. The invention further extends to the use of said sequences as signal sequences. The invention further extends to the translation product of nucleotides 53 to 136 in SEQ ID No. 1, nucleotides 11 to 94 in SEQ ID No.2, nucleotides 24 to 107 in SEQ ID No. 3, nucleotides 20 to 103 in SEQ ID No. 4 or nucleotides 1 to 176 in SEQ ID No. 5 and protein having an amino acid sequence which is at least 85% similar to said product.

It is preferred that the degree of similarity is at least 90%, more preferred that the degree of similarity is at least 95% and still more preferred that the degree of similarity is at least 97%.

In the context of the present invention, two amino acid sequences with at least 85% similarity to each other have at least 85% similar (identical or conservatively replaced) amino acid residues in a like position when aligned optimally allowing for up to 3 gaps, with the *proviso* that in respect of the gaps a total of not more than 15 amino acid residues is affected. Likewise, two amino acid sequences with at least 90% similarity to each other have at least 90% identical or conservatively replaced amino acid residues in a like position when aligned optimally allowing for up to 3 gaps with the *proviso* that in respect of the gaps a total of not more than 15 amino acid residues is affected.

For the purpose of the present invention, a conservative amino acid is defined as one which does not alter the activity/function of the protein when compared with the unmodified protein. In particular, conservative replacements may be made between amino acids within the following groups:

(i) Alanine, Serine, Glycine and Threonine

(ii) Glutamic acid and Aspartic acid

(iii) Arginine and Lysine

(iv) Isoleucine, Leucine, Valine and Methionine

(v) Phenylalanine, Tyrosine and Tryptophan

The invention also includes a polynucleotide encoding a protein having a substantially similar activity to any one of those encoded by SEQ ID No. 1 to SEQ ID No. 5,

which polynucleotide is complementary to one which when incubated at a temperature of between 55 and 65°C in 0.3 strength citrate buffered saline containing 0.1% SDS followed by rinsing at the same temperature with 0.1 strength citrate buffered saline containing 0.1% SDS still hybridises with a sequence depicted in SEQ ID No. 1 to SEQ ID No.5. with the proviso that said sequence is not that described in SEQ ID No. 6 or SEQ ID No. 7.

The polynucleotide sequence provided in SEQ ID Nos 6 and 7 is the predicted DNA sequence for Dm-AMP1 and Dm-AMP2 as described in Figure 31A of Published International Patent Application No. WO 93/05153.

The invention still further includes a polynucleotide encoding a protein having a substantially similar activity to that encoded by nucleotides 137 to 286 in SEQ ID No. 1, by nucleotides 95 to 244 in SEQ ID No. 2, by nucleotides 108 to 257 in SEQ ID No. 3, by nucleotides 104 to 253 in SEQ ID No. 4 or by nucleotides 177 to 326 in SEQ ID No.5 which polynucleotide is complementary to one which when incubated at a temperature of between 55 and 65°C in half strength citrate buffered saline containing 0.1% SDS followed by rinsing at the same temperature with 0.3 strength citrate buffered saline containing 0.1% SDS still hybridises with the sequence comprised by nucleotides 137 to 286 in SEQ ID No. 1, by nucleotides 95 to 244 in SEQ ID No. 2, by nucleotides 108 to 257 in SEQ ID No. 3, by nucleotides 104 to 253 in SEQ ID No. 4 or by nucleotides 177 to 326 in SEQ ID No.5.with the proviso that said sequence is not that described in SEQ ID No. 6 or SEQ ID No. 7.

When the test and inventive sequences (SEQ ID Nos 1 to 5) are double stranded the nucleic acid constituting the test sequence preferably has a TM within 15°C of that of the said inventive sequence. In the case that the test and inventive sequences are mixed together and are denatured simultaneously, the TM values of the sequences are preferably within 5°C of each other. More preferably the hybridisation is performed under stringent conditions, with either the test or inventive sequences preferably being supported. Thus either a denatured test or any of the inventive sequences is preferably first bound to a support and hybridisation is effected for a specified period of time at a temperature of between 55 and 65°C in 0.3 strength citrate buffered saline containing 0.1% SDS followed by rinsing of the support at the same temperature but with 0.1 strength citrate buffered saline. Where the hybridisation involves a fragment of the inventive sequence, the hybridisation conditions may be less stringent, as is indicated in the immediately preceding paragraph.

It may be desired to target the translation products of the polynucleotide to specific sub-cellular compartments within the plant cell, in which case the polynucleotide comprises sequences encoding chloroplast transit peptides, cell wall targeting sequences etc. immediately 5' of the regions encoding the said translation products.

Translational expression of the protein encoding sequences contained within the polynucleotide may be relatively enhanced by including known non translatable translational enhancing sequences 5' of the said protein encoding sequences. The skilled man is very familiar with such enhancing sequences, which include the TMV-derived sequences known as omega, and omega prime, as well as other sequences derivable, *inter alia*, from the regions 5' of other viral coat protein encoding sequences.

In a particularly preferred embodiment of the invention, the polynucleotide is modified in that mRNA instability motifs and/or fortuitous splice regions are removed, or plant preferred codons are used so that expression of the thus modified polynucleotide in a plant yields substantially similar protein having a substantially similar activity/function to that obtained by expression of the unmodified polynucleotide in the organism in which the protein encoding regions of the unmodified polynucleotide are endogenous, with the *proviso* that if the thus modified polynucleotide comprises plant preferred codons, the degree of identity between the modified polynucleotide and a polynucleotide endogenously contained within the said plant and encoding substantially the same protein is less than about 60%.

The invention also includes a plant transformation vector comprising a plant operable promoter, a polynucleotide sequence comprising all or part of the sequence selected from those depicted in SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 3, SEQ ID No. 4 and SEQ ID No.5 under the transcriptional control thereof and encoding an antimicrobial protein, and a plant operable transcription terminator. The promoter may be constitutive or inducible. In particular, the promoter may be such that it induces transcription in response to application to the plant material containing it of a chemical.

The polynucleotide sequences provided in SEQ ID Nos 1 to 5 are related sequences with the translated products thereof showing a high degree of sequence similarity and it is believed that they may belong to a multi gene family.

The invention still further includes plant tissue transformed with the said polynucleotide or vector, and material derived from the said transformed plant tissue, as well

as morphologically normal fertile whole plants comprising the tissue or material. Such transformed plants include but are not limited to, field crops, fruits and vegetables such as canola, sunflower, tobacco, sugar beet, cotton, maize, wheat, barley, rice, sorghum, tomato, mango, peach, apple, pear, strawberry, banana, melon, potato, carrot, lettuce, cabbage, onion, etc. Particularly preferred genetically modified plants are soya spp, sugar cane, pea, field beans, poplar, grape, citrus, alfalfa, rye, oats, turf and forage grasses, flax and oilseed rape, and nut producing plants insofar as they are not already specifically mentioned .

The invention still further includes the progeny of the plants of the preceding paragraph, which progeny comprises the polynucleotide of the invention stably incorporated into its genome and heritable in a mendelian manner and the seeds of such plants and such progeny.

The invention also provides a method of producing plants which are substantially tolerant or substantially resistant to antimicrobial infection, comprising the steps of:

(i) transforming plant material with the polynucleotide or vector of the invention;

(ii) selecting the thus transformed material; and

(iii) regenerating the thus selected material into morphologically normal fertile whole plants.

Plant transformation, selection and regeneration techniques, which may require routine modification in respect of a particular plant species, are well known to the skilled man.

The invention will be further apparent from the following description taken in conjunction with the associated figures and sequence listings in which :

Figure 1 shows the polynucleotide and corresponding amino acid sequences for A) Dm1 (SEQ ID No5) and B) Dm2.18 (SEQ ID No1),

Figure 2 shows the polynucleotide and corresponding amino acid sequences for A) Dm2.1 (SEQ ID No2) and B) Dm2.3 (SEQ ID No3),

Figure 3 shows the polynucleotide and corresponding amino acid sequence for Dm2.5 (SEQ ID No. 4),

Figure 4 shows a diagrammatic map of plasmids pMJB1, pDmAMPD and pDmAMPE;

Figure 5 shows a diagrammatic map of plasmid pFAJ3106;

Figure 6 shows a diagrammatic map of plasmid pFAJ3109

Figure 7 shows the nucleotide sequence between the XhoI and SacI sites of pFAJ3106;

Figure 8 shows the nucleotide sequence between the XhoI and SacI sites of pFAJ3109;

Figure 9 shows a diagrammatic map of plasmid pZPS38;

Figure 10 shows a diagrammatic map of plasmid pZPS34;

Figure 11 shows a diagrammatic map of plasmid pZPS35;

Figure 12 shows a diagrammatic map of plasmid pZPS37.

Figure 13 shows a plan of the construction of the Dm-AMP gene

Figure 14 shows the polynucleotide sequence for DmAMP1 (SEQ ID No. 6) and Dm-AMP2 (SEQ ID No. 7).

**EXAMPLE 1**

**Dm Gene Isolation and Vector Construction**
**Dahlia cDNA library construction**

Near-dry seeds were collected from flowers of *Dahlia merkii* .

Total RNA was purified from the seeds using the method of Jepson *et al* .

Seeds were frozen in liquid nitrogen and ground to a fine powder using a mortar and pestle. Phenol/*m*-cresol (9:1) was added followed by RNA homogenisation buffer and the mixture ground until a fine paste was obtained. The mixture was spun, the aqueous phase collected and extracted twice with phenol/chloroform (1:1). Lithium chloride (12 M) was added to the resulting aqueous layer to a final concentration of 2 M and incubated overnight at 4°C. Precipitated RNA was collected by spinning at 13,000 rpm in an Eppendorf centrifuge and the RNA pellet re-suspended in 5 mM Tris-HCl, pH 7.5. A second overnight lithium chloride precipitation was carried out and the RNA collected and re-suspended in 5 mM Tris-HCl, pH 7.5.

0.6 mg of total RNA was obtained from 2 g of *D. merkii* seed.

PolyATract magnetic beads (Promega) were used to isolate approximately 2 μg poly-$A^+$ RNA from 0.2 mg of total RNA.

The poly-$A^+$ RNA was used to construct a cDNA library using a ZAP-cDNA synthesis kit (Stratagene). Following first and second strand synthesis double stranded cDNA was size fractionated on a Sephacryl S-400 column. The three largest cDNA size fractions were pooled and ligated with vector DNA. After phage assembly using Gigapack Gold (Stratagene) packaging extracts, approximately 1 x $10^5$ pfu were obtained.

**Probing**

Using oligos AFP-5 (based on n-ter sequence CEKASKTW) and AFP-3EX (based on C-ter sequence MCFCYFNC) and Dahlia genomic DNA, a 144 bp PCR product was produced and isolated from an agarose gel. The PCR product was cloned into pBluescript. The insert of 10 transformants were sequenced. The sequences represented 3 Dm-AMP1 genes one of which, PCR clone 4, encoded the observed mature Dm-AMP1.

The 144 bp PCR product mixture labelled with $\alpha^{32}$-P d-CTP was used to probe Hybond N (Amersham) filter lifts made from plates containing a total of 6 x $10^5$ pfu of the cDNA library. The filters were hybridised at 46°C for 18 hrs in 5 x SSPE, 0.01% SDS, 0.25% skimmed milk powder. Filters were washed in 2 x SSC, 0.1% SDS at 60°C. Autoradiography was carried out at -70°C with intensifying screens. Thirty potentially positive signals were observed. 22 plaques were picked and taken through two further rounds of screening. After in vivo excision 13 clones were characterised by DNA sequencing.

Four classes of Dm-AMP related peptide were encoded by the 13 cDNA clones and the sequences of these peptides are provided in SEQ ID Nos 1- 4 in the accompanying figures. Three versions of the Dm-AMP core region were represented in the four classes. One of the classes (Dm2.5 type) contained a core region which may correspond to Dm-AMP2.

None of the cDNAs encoded a core region equivalent to the observed mature Dm-AMP1 peptide sequence.

**Isolation of a Mature Dm-AMP1 Gene**

Using the sequence of PCR clone 4 (above) and information from the $NH_2$ and COOH ends of the peptides described by cDNA sequences two pairs of oligonucleotides were designed for amplification of a gene encoding the observed mature Dm-AMP1.

Dahlia genomic DNA was PCRed with MATAFP-5P and MATAFP-5, the PCR product was cloned into pBluescript and clones were sequenced. A clone containing the 5' half of a Dm-AMP1 gene was identified.

Dahlia genomic DNA was PCRed with MATAFP-3 and DM25MAT-3, the PCR product was cloned into pBluescript and clones were sequenced. A clone containing the 3' half of a Dm-AMP1 gene was identified.

The 5' and 3' sections of the mature gene were combined to assemble the sequence of the mature Dm-AMP1 gene (see below) . Exon 1, 64 bp encoding part of the leader peptide, 92

bp intron and exon 2 encoding the end of the leader sequence, Dm-AMP1 core and C-terminal extension.

**Vector Oligonucleotide design**

Four oligonucleotides were designed based on the DNA sequence of the mature Dm-AMP1 gene:

DMVEC-1 top strand priming at the 5' end of the mature DmAMP-1 gene incorporating a Nco I site at the translation start of DmAMP-1 allowing cloning into pMJB1 (see Figure 4).

DMVEC-2 bottom strand priming in the 3' end of the C-terminal extension and a Sac I site for cloning in pMJB1.

DMVEC-3 top strand priming at the 5' end of the mature DmAMP-1 gene incorporating a Nco I site at the translation start of DmAMP-1 allowing cloning into pMJB1 also encoding complete signal peptide (minus intron).

DMVEC-4 bottom strand priming in the 3' end of the core region and a Sac I site for cloning in pMJB1.

Dahlia genomic DNA was used as template in a PCR with oligonucleotides DMVEC-1 and DMVEC-2. A product of approximately 450 bp was obtained, this PCR product was used as a template for PCRs described in vector construction below.

**Vector Construction**

**pDmAMPD**

The PCR product obtained with DMVEC-1 and DMVEC-4 was cut with Nco I and Sac I the 60 bp Nco I/Sac I fragment was isolated and cloned into pMJB1cut with Nco I and Sac I. The fragment in one of the resulting transformants was confirmed by sequencing, the clone was termed pDmAMPA.

The PCR product obtained with DMVEC-3 and DMVEC-4 was cut with Nco I, the 150 bp Nco I fragment isolated and cloned into pDmAMPA cut with Nco I. The region of DNA encoding Dm-AMP leader plus core region in one of the resulting transformants was sequenced (see sequence below), the clone was termed pDmAMPD.

**pDmAMPE**

The PCR product obtained with DMVEC-1 and DMVEC-2 was cut with Nco I and Sac I the 180 bp Nco I/Sac I fragment was isolated and cloned into pMJB1cut with Nco I and Sac I.

The fragment in one of the resulting transformants was confirmed by sequencing, the clone was termed pDmAMPB.

The PCR product obtained with DMVEC-3 and DMVEC-4 was cut with Nco I, the 150 bp Nco I fragment isolated and cloned into pDmAMPB cut with Nco I. The region of DNA encoding Dm-AMP leader, core plus C-terminal extension in one of the resulting transformants was sequenced (see sequence below), the clone was termed pDmAMPE. Both vector sequences contain PCR derived base changes with respect to the Dm-AMP1 gene however the base changes do not result in amino acid substitutions.

AFP-5 (to CEKASKTW)

TG(T,C)GANAANGCN(A,T)(G,C)NAA(A,G)ACNTGG

AFP-3EX (to MCFCYFNC)

CA(A,G)TT(A,G)AANTANCANAAA(A,G)CACAT

MATAFP-5P

ATGGC(C,G)AAN(A,C)(A,G)NTC(A,G)GTTGCNTT

MATAFP-5

AAACACATGTGTTTCCCATT

MATAFP-3

AGCGTGTCATGTGCGTAAT

Dm25MAT-3

TAAAGAAACCGACCCTTTCACGG

DMVEC-1

ATCGTAGCCATGGTGAATCGGTCGGTTGCGTTCTCCGCG

DMVEC-2

AAACCGACCGAGCTCACGGATGTTCAACGTTTGGAAC

DMVEC-3

ATGCATCCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTTTTCGTGCTC
GCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATGCGAGAAA

DMVEC-4

AGCAAGCTTTTCGGGAGCTCAACAATTGAAGTAA

**EXAMPLE 2**

**Constructions of plant transformation vectors**

Expression cassettes containing a Dm-AMP1 open reading frame functionally linked to an enhanced 35S promoter, a TMV omega translational enhancer and a Nos 3' region are

isolated as restriction fragments. pDmAMPD and pDmAMPE are both digested with the restriction endonucleases HindIII and EcoRI and the appropriate restriction fragment isolated and purified. Each fragment is ligated into a binary vector (a pBIN19 derivative named pBin19i) which has also been digested with HindIII and EcoRI. The resulting constructs, named pDmAMPLC and pDmAMPLCC, incorporate the expression cassettes from pDmAMPD and pDmAMPE respectively.

pDmAMPLC and pDmAMPLCC are subsequently introduced into *Agrobacterium tumefaciens* strain LBA4404 and introduced into tobacco and oil seed rape using standard plant transformation methodology.

Plants are regenerated from callus tissue resistant to the selective agent kanamycin and expression of the Dm-AMP1 product is monitored by standard Western blot or ELISA methods using antibody which had been raised against Dm-AMP1 protein. A range of expression levels are detected. The Dm-AMP1 expressed in selected transgenic is further characterised following extraction and partial purification from leaves of such lines. The product is of the predicted mass, as indicted by mass spectrometry. It is also demonstrated to retain biological activity after extraction as demonstrated by retention of antifungal activity in in-vitro (micro-titre plate) assays.

**EXAMPLE 3**

**Constructions of plant transformation vectors for polyprotein expression**

Schematic representations of the plant transformation vectors used in this work, pFAJ3106 and pFAJ3109, are shown in figures 5 and 6, respectively. The nucleotide sequences comprised between the *Xho*I and *Sac*I sites of these plasmids, which encompass the regions encoding antimicrobial proteins, are presented in Figures 7 and 8. The regions comprised between the *Xho*I and *Sac*I sites of plasmid pFAJ3106 (shown in Figure 7) was constructed following the two-step recombinant PCR protocol of Pont-Kindom G.A.D. (1994, Biotechniques 16, 1010-1011). Primers OWB175 (5'AGGAAGTTCATTTCATTTGG) and OWB279 (5'-GCCTTTGGCACAACTTCTGCCTCTTTCCGATGAGTTGTTCGGCTTTAAGTTTGTC); were used in a first PCR reaction with plasmid pDMAMPE (see above) as a template. The second PCR reaction was done using as a template plasmid pFRG4 (Terras F.R.G. *et al.*, 1995, Plant Cell 7, 573-588) and as primers a mixture of the PCR product of the first PCR reaction, primer OWB175 and primer OWB172 (5'TTAGAGCTCCTATTAACAAGGAAAGTAGC,

*Sac*I site underlined). The resulting PCR product was digested with *Xho*I and *Sac*I and cloned into the expression cassette vector pMJB1 (see above). The expression cassette in the resulting plasmid, called pFAJ3099, was digested with *Hin*dIII (flanking the 5` end of the CaMV35S promoter) and *Eco*RI (flanking the 3` end of the nopaline synthase terminator) and cloned in the corresponding sites of the plant transformation vector pGPTVbar (Becker D. *et al.*, 1992, Plant Mol. Biol. 20, 1195-1197) to yield plasmid pFAJ3106.

Plasmid pFAJ3109 was constructed by cloning the *Hin*dIII-*Eco*RI fragment of plasmid pDMAMPD (see above) into the corresponding sites of plant transformation vector pGPTVbar (see above).

**Plant transformation**

*Arabidopsis thaliana* ecotype Columbia-O was transformed using recombinant *Agrobacterium tumefaciens* by the inflorescence infiltration method of Bechtold N. *et al.* (1993, C.R. Acad. Sci. 316, 1194-1199). Transformants were selected on a sand/perlite mixture subirrigated with water containing the herbicide Basta (Agrevo) at a final concentration of 5 mg/l for the active ingredient phosphinothricin.

**Elisa assays and protein assays**

Antisera were raised in rabbits injected with either RsAFP2 (purified as described in Terras F.R.G. *et al.*, 1992, J. Biol. Chem. 267, 15301-15309) or DmAMP1 (purified as in Osborn R.W. *et al.*, 1995, FEBS Lett. 368, 257-262). ELISA assays were set up as competitive type assays essentially as described by Penninckx I.A.M.A. *et al.* (1996, Plant Cell 8, 2309-2323). Coating of the ELISA microtiter plates was done with 50 ng/ml RsAFP2 or DmAMP1 in coating buffer. Primary antisera were used as 1000- and 2000-fold diluted solutions (DmAMP1 and RsAFP2, respectively) in 3 % (w/v) gelatin in PBS containing 0.05 % (v/v) Tween 20.

Total protein content was determined according to Bradford (1976, Anal. Biochem. 72, 248-254) using bovine serum albumin as a standard.

**Purification and characterisation of expressed proteins**

Arabidopsis leaves were homogenized under liquid nitrogen and extracted with a buffer consisting of 10 mM $NaH_2PO_4$, 15 mM $Na_2HPO_4$, 100 mM KCl, 1.5 M NaCl. The homogenate was heated for 10 min at 85°C and cooled down on ice. The heat-treated extract was centrifuged for 15 min at 15 000 x *g* and was injected on a reserved phase high pressure

liquid chromatography column (RP-HPLC) consisting of C8 silica (0,46 cm x 25 cm; Rainin) equilibrated with 0.1 % (v/v) trifluoroacetic acid (TFA). The column was eluted at 1 ml/min in a linear gradient in 35 min. from 15 % to 50 % (v/v) acetonitrile in 0.1 % (v/v) TFA. The eluate was monitored for absorbance at 214 nm, collected as 1 ml fractions, evaporated and finally redissolved in water. The fractions were tested by ELISA assays.

**Preparation of extracellular fluid and intracellular extract**

Intercellular fluid was collected from Arabidopsis leaves by immersing the leaves in a beaker containing extraction buffer (10 mM $NaH_2PO_4$, 15 mM $Na_2HPO_4$, 100 mM KCl, 1.5 M NaCl). The beaker with the leaves was placed in a vacuum chamber and subjected to six consecutive rounds of vacuum for 2 min followed by abrupt release of vacuum. The infiltrated leaves were gently placed in a centrifuge tube on a grid separated from the tube bottom. The intercellular fluid was collected from the bottom after centrifugation of the tubes for 15 min at 1800 x *g*. The leaves were resubjected to a second round of vacuum infiltration and centrifugation and the resulting (extracellular) fluid was combined with that obtained after the first vacuum infiltration. After this step the leaves were extracted in a Phastprep (BIO101/Savant) reciprocal shaker and the extract clarified by centrifugation (10 min at 10,000 x *g*) and the resulting supernatant considered as the intracellular extract.

**Characterization of transgenic plants and expression analysis**

To explore the possibility of expressing polyprotein precursor genes in plants, three different plant transformation vectors were made with the aim to co-express two different cysteine-rich plant defensins with antifungal properties, namely RsAFP2 and DmAMP1. The polyprotein precursor regions of these constructs all featured a leader peptide region from the DmAMP1 cDNA, the mature protein domain of DmAMP1, an internal propeptide region, and the mature protein domain of RsAFP2. The three constructs differed only in the internal propeptides

- construct 3106 has a propeptide consisting of a part of the DmAMP1 propeptide and a putative subtilisin-like protease processing site (IGKR) at its C-terminus.

The rationale behind construct 3106, is based on our observations that the C-terminal propeptides of DmAMP1 are cleaved off at their N-terminus when expressed as DmAMP1-preproproteins in tobacco, respectively, while this processing event does not detract the mature proteins from being sorted to the apoplast (De Bolle *et al.*, 1996, Plant Mol. Biol. 31, 993-1008; R.W. Osborn and S. Attenborough, personal communication). This infers that the

processing enzymes are either in the secretory pathway or in the apoplast. On the other hand, C-terminal cleavage of the internal propeptide in these constructs should be executed by a subtilisin-like protease, a member of which in yeast (Kex2) is known to occur in the Golgi apparatus (Wilcox C.A. and Fuller R.S., 1991, J. Cell. Biol. 115, 297- ), while a member in tomato occurs in the apoplast (Tornero P. *et al.*, 1997, J. Biol. Chem. 272, 14412-14419). Proteins deposited in the apoplast, the preferred deposition site for antimicrobial proteins engineered in transgenic plants (Jongedijk E. *et al.*, 1995, Euphytica 85, 173-180; De Bolle *et al.*, 1996, Plant Mol. Biol. 31, 993-1008) are normally synthesized via the secretory pathway, encompassing the Golgi apparatus.

Constructs were also made for expression of DmAMP1 (construct 3109, figure 6). Expression levels of DmAMP1 and RsAFP2 were analysed in leaves taken from a series of T1 transgenic Arabidopsis plants resulting from transformation with the constructs described above. The results of the expression analyses based on Elisa assays are presented in Table 1. Most of the tested lines transformed with the polyprotein constructs 3106 clearly expressed both DmAMP1 and RsAFP2. There was generally a good correlation between DmAMP1 and RsAFP2 levels. However, the RsAFP2 levels were generally 2 to 5-fold lower than the DmAMP1 levels. It is not known whether the apparent lower expression levels of RsAFP2 versus DmAMP1 are real or whether they result from a bias in the extraction procedure or the assays. The expression levels in the lines transformed with the polyprotein constructs 3106 were generally much higher compared to those in lines transformed with the single protein construct 3109. Hence, the use of polyprotein constructs appears to result in enhanced expression, which is an unexpected finding.

**Analysis of the proteins expressed for polyprotein constructs**

A transgenic line was selected among each of the populations transformed with either construct 3106 and the selected lines were further bred to obtain plants homozygous for the transgenes. In order to analyse whether DmAMP1 and RsAFP2 were correctly processed in these lines, extracts from the plants were prepared as described in Materials and Methods and separated by RP-HPLC on a C8-silica column. Fractions were collected and assessed for presence of compounds cross-reacting with antibodies raised against either DmAMP1 or RsAFP2 using Elisa assays.

DmAMP1 cross-reacting compound eluted at a position identical or very close to that of authentic DmAMP1 in the line transformed with construct 3106. Likewise, a RsAFP2 cross-reacting compound was detected in the 3106 lines at an elution position identical or very close to that of authentic RsAFP2. None of the fractions reacted with both the anti-DmAMP1 and anti-RsAFP2 antibodies, indicating that an uncleaved fusion protein was not present in the extracts. No cross-reacting compounds were observed in a non-transformed line.

It is concluded that the primary translation products of the transcription units of construct 3106 (partial DmAMP1 C-terminal propeptide with subtilisin-like protease site as a linker peptide) are somehow processed to yield separate DmAMP1-cross-reacting and RsAFP2-cross-reacting portions that appear to be identical or very closely related to DmAMP1 and RsAFP2, respectively, based on their chromatographic behavior.

**Analysis of the subcellular location of coexpressed plant defensins**

In order to determine whether the coexpressed plant defensins are either secreted extracellularly or deposited intracellularly, extracellular fluid and intracellular extract fractions were obtained from leaves of homozygous transgenic Arabidopsis lines transformed with constructs 3106. The cytosolic enzyme glucose-6-phosphate dehydrogenase was used as a marker to detect contamination of the extracellular fluid fraction with intracellular components. As shown in Table 1, glucose-6-phosphate dehydrogenase was partitioned in a ratio of about 80/20 between intracellular extract fractions and extracellular fluid fractions. In contrast, the majority of DmAMP1 and RsAFP2 content in all transgenic plants tested was found in the extracellular fluid fractions. These results indicate that both plant defensins released from the polyprotein precursors are deposited primarily in the apoplast. Hence, all processing steps that result in cleavage of the polyprotein structure must occur either in the apoplast or along the secretory pathway.

Table 1: Relative abundance of glucose-6-phosphate dehydrogenase activity (GPD), DmAMP1 and RsAFP2 in the extracellular fluid (EF) and intracellular extract (IE) fractions obtained from transgenic Arabidopsis plants.

| Construct | Relative abundance[1] (%) of | | | | | |
|---|---|---|---|---|---|---|
| | GPD | | DmAMP1 | | RsAFP2 | |
| | EF | IE | EF | IE | EF | IE |
| pFAJ3106 | 17 | 83 | 94 | 6 | 60 | 40 |

Relative abundance is expressed as % of the sum of the contents in the EF and IE fractions.

**EXAMPLE 4**

**Expression of the sweet tasting protein Brazzein in tomato**

Production of transgenic tomato plants with increased accumulation of sweet tasting protein Brazzein.

Constructs were prepared containing the Dahlia (*Dahlia merckii*) antimicrobial protein signal peptide fused with the Brazzein under the transcriptional control of the *Arabidopsis* polyubiquitin extension protein promoter (UBQ) or the Polygalacturonase promoter (PG). Constructs were also prepared which encoded Brazzein without a signal peptide but with an N-terminal methionine by the insertion of ATG nucleotides upstream of the Brazzein gene under the expressional control of either the UBQ promoter or the PG promoter. These were prepared as follows:

Construction of the transformation vector for expression in tomato with Dahlia signal peptide fused to Brazzein under the expressional control of either the UBQ promoter or the PG promoter:

A synthetic DNA was produced which coded for the with the Dahlia signal peptide fused to Brazzein. The codons were optimised for expression in tomato. Using appropriate restriction sites the coding sequence was cloned into a plasmid vector. The coding region was excised from the plasmid and cloned between the promoter in question and the terminator in the correct orientation for expression.

**Generation and analysis of plants transformed with the transformation vector.**

The vector was transferred to *Agrobacterium tumefaciens* LBA4404 (a microorganism widely available to plant biotechnologists) and used to transform tomato plants. Transformation of tomato stem segments followed standard protocols (e.g. Bird *et al* Plant Molecular Biology 11, 651-662, 1988). Transformed plants were identified by their ability to grow on media containing the antibiotic kanamycin. Up to 30 individual plants were regenerated with each construct and grown to maturity. The presence of the construct in all of the plants was confirmed by polymerase chain reaction analysis. DNA Southern blot analysis on all plants indicated that the insert copy number was between 1 and 10. Northern blot analysis on fruit from one plant indicated that the Brazzein gene was expressed.

Brazzein production in the fruit of all plants was measured by ELISA (enzyme linked imunoabsorption assay) using a polyclonal and a monoclonal antibody raised against native Brazzein protein isolated from the fruit of the plant *Pentadiplandra brazzeana Baillon* Two fruit were collected from each transgenic plant at 7 days post breaker (the term breaker is used to indicate when the tomato fruit first show signs of the orange colouration characteristic of most mature tomato fruit). Total fruit protein was extracted from a sample of the pericarp of each of the fruit. The amount of Brazzein protein in the total protein extract was measured by ELISA and calculated as the amount of Brazzein per gram fresh weight of the fruit. For each plant the average Brazzein content of the two fruits was calculated. In some plants Brazzein could not be detected in the fruit using the ELISA technique. Western blot analysis of the total protein extract from some of the fruit revealed a 6.5kD protein band. which matches the predicted size of the mature Brazzein protein. This confirmed that the fruit contained Brazzein and that the signal peptide had been cleaved as if the signal peptide had not been cleaved, one would expect the protein to be larger. The Brazzein in fruit from plants which had been transformed with a construct lacking a signal peptide was not detected by Western blot. This is because the Brazzein content in these fruit is below the level of detection by western blot. ELISA is a more sensitive technique than western blot and protein was detected in these fruit by this method.

The results are summarised in Table 2 below.

TABLE 2

| Construct Name | Promoter | Signal Peptide | No. of Plants Tested | Plants expressing Brazzein | Max Brazzein ng/g Fresh wt | Min Brazzein ng/g Fresh wt |
|---|---|---|---|---|---|---|
| pZPS34 | UBQ | None | 29 | 18 | 25.57 | Not Detected |
| pZPS35 | UBQ | Dahlia AMP1 | 25 | 23 | 226.53 | Not Detected |
| pZPS37 | PG | None | 15 | 7 | 12.77 | Not Detected |
| pZPS38 | PG | Dahlia AMP1 | 13 | 11 | 51745.77 | Not Detected |

## CLAIMS

1. A polynucleotide comprising a sequence selected from those depicted in SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 3, SEQ ID No. 4 and SEQ ID No. 5.

2. A polynucleotide comprising nucleotides 137 to 286 in SEQ ID No. 1, nucleotides 95 to 244 in SEQ ID No. 2, nucleotides 108 to 257 in SEQ ID No. 3, nucleotides 104 to 253 in SEQ ID No. 4 or nucleotides 177 to 326 in SEQ ID No. 5.

3. A polynucleotide comprising nucleotides 287 to 385 in SEQ ID NO. 1, nucleotides 245 to 334 in SEQ ID No. 2, nucleotides 258 to 317 in SEQ ID No. 3, nucleotides 254 to 343 in SEQ ID No. 4 or nucleotides 327 to 446 in SEQ ID No.5.

4. A polynucleotide comprising nucleotides 53 to 136 in SEQ ID No. 1, nucleotides 11 to 94 in SEQ ID No.2, nucleotides 24 to 107 in SEQ ID No. 3, nucleotides 20 to 103 in SEQ ID No. 4 or nucleotides 1 to 176 in SEQ ID No. 5.

5. A polynucleotide according to claim 4 comprising nucleotides 1 to 176 in SEQ ID No. 5 excluding the sequence encoding the intron marked at positions 65 to 156.

6. A polynucleotide encoding a protein having a substantially similar activity to that encoded by any of SEQ ID No. 1, No 2, No. 3, No,4 or No. 5 which polynucleotide is complementary to one which when incubated at a temperature of between 55 and 65°C in 0.3 strength citrate buffered saline containing 0.1% SDS followed by rinsing at the same temperature with 0.1 strength citrate buffered saline containing 0.1% SDS still hybridises with the sequence depicted in SEQ ID No. 1, or No 2 or No. 3, or No. 4 or No.5. with the proviso that said sequence is not that in SEQ ID No. 6 or SEQ ID No. 7.

7. A polynucleotide encoding a protein having a substantially similar activity to that encoded by nucleotides 137 to 286 in SEQ ID No. 1, nucleotides 95 to 244 in SEQ ID No. 2, nucleotides 108 to 257 in SEQ ID No. 3, nucleotides 104 to 253 in SEQ ID No. 4 or nucleotides 177 to 326 in SEQ ID No. 5., which polynucleotide is complementary to one which when incubated at a temperature of between 55 and 65°C in half strength citrate buffered saline containing 0.1% SDS followed by rinsing at the same temperature with 0.3 strength citrate buffered saline containing 0.1% SDS still hybridises with the sequence comprised by nucleotides 137 to 286 in SEQ ID No. 1, nucleotides 95 to 244 in SEQ ID No. 2, nucleotides 108 to 257 in SEQ ID No. 3, nucleotides 104 to 253 in SEQ ID No. 4 or nucleotides 177 to 326 in SEQ ID No. 5. with the proviso that said sequence is not that in SEQ ID No. 6 or SEQ ID No. 7.

8. A polynucleotide according to any preceding claim, further comprising a region encoding a peptide which is capable of targeting the translation products of the sequence to plastids such as chloroplasts, mitochondria, other organelles or plant cell walls.

9. A polynucleotide according to any preceding claim, wherein translational enhancing sequences are inserted 5' of the protein encoding regions comprised by the polynucleotide.

10. A polynucleotide according to any preceding claim, which is modified in that mRNA instability motifs and/or fortuitous splice regions are removed, or plant preferred codons are used so that expression of the thus modified polynucleotide in a plant yields substantially similar protein having a substantially similar activity/function to that obtained by expression of the unmodified polynucleotide in the organism in which the protein encoding regions of the unmodified polynucleotide are endogenous, with the *proviso* that if the thus modified polynucleotide comprises plant preferred codons, the degree of identity between the modified polynucleotide and a polynucleotide endogenously contained within the said plant and encoding substantially the same protein is less than about 60%.

11. A plant transformation vector comprising a plant operable promoter, a polynucleotide sequence according to claims 1 to 10 under the transcriptional control thereof and a plant transcription terminator.

12. Plant tissue transformed with the polynucleotide of any one of claims 1 to 10 or the vector of claims 11 and material derived from the said transformed plant tissue.

13. Morphologically normal fertile whole plants comprising the tissue or material of the preceding claim.

14. The progeny of the plants of the preceding claim, which progeny comprises the polynucleotide of any one of claims 1 to 10 stably incorporated into its genome and heritable in a mendelian manner, the seeds of such plants and such progeny.

15. A method of producing plants which are substantially tolerant or substantially resistant to microbial infection, comprising the steps of:

(i) transforming plant material with the polynucleotide of any one of claims 1 to 10 or the vector of claim 11

(ii) selecting the thus transformed material; and

(iii) regenerating the thus selected material into morphologically normal fertile whole plants.

16. Use of the polynucleotide of any one of claims 1 to 10, or the vector of claim 11 in the production of plant tissues and/or morphologically normal fertile whole plants which are substantially tolerant or substantially resistant to microbial infection.

17. The translation product of the region comprised by nucleotides 137 to 286 in SEQ ID No. 1.; nucleotides 95 to 244 in SEQ ID No.2, nucleotides 108 to 257 in SEQ ID No. 3 and nucleotides 104 to 253 in SEQ ID No. 4. and protein having an amino acid sequence which is at least 95% similar to said product.

18. The translation product of the region comprised by nucleotides 287 to 385 in SEQ ID NO. 1, nucleotides 245 to 334 in SEQ ID No. 2, nucleotides 258 to 317 in SEQ ID No. 3, nucleotides 254 to 343 in SEQ ID No. 4 or nucleotides 327 to 446 in SEQ ID No.5 and protein having an amino acid sequence which is at least 85% similar to said product.

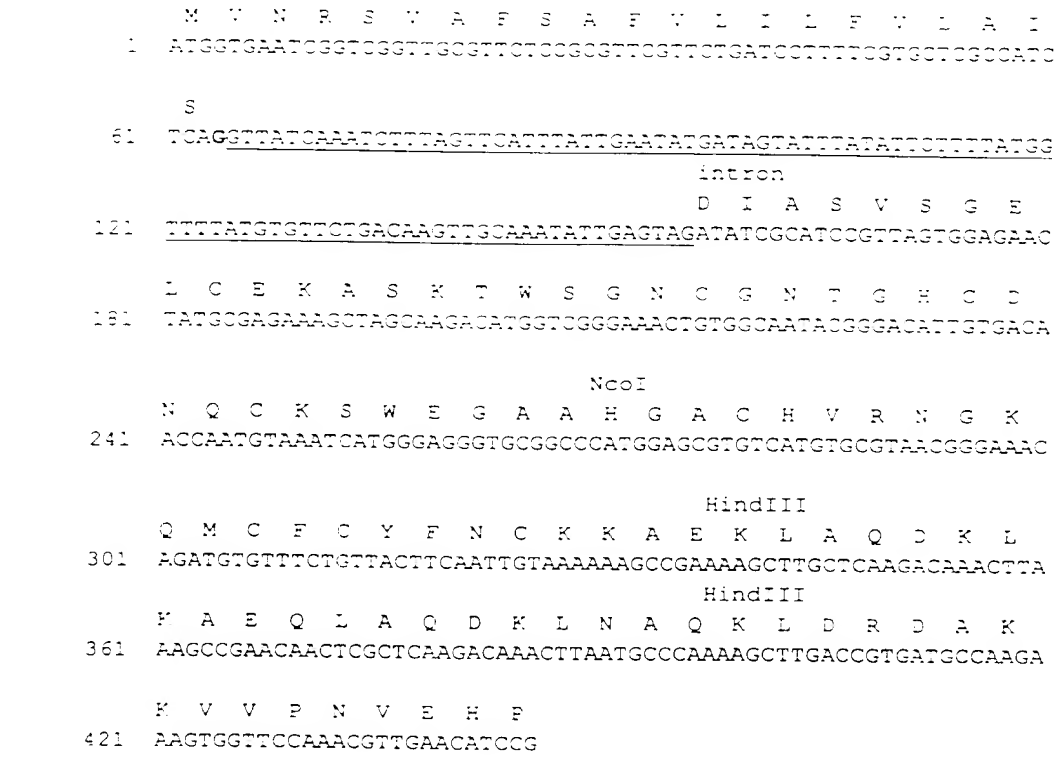
19. The translation product of the region comprised by nucleotides 53 to 136 in SEQ ID No. 1, nucleotides 11 to 94 in SEQ ID No.2, nucleotides 24 to 107 in SEQ ID No. 3, nucleotides 20 to 103 in SEQ ID No. 4 or nucleotides 1 to 176 in SEQ ID No. 5 and protein having an amino acid sequence which is at least 85% similar to said product.

20. A method of selectively controlling microorganisms at a locus comprising the plants, progeny and/or seeds of either of claims 13 or 14, comprising applying to the locus a microorganism controlling amount of the translation product of the region comprised by nucleotides 137 to 286 in SEQ ID No. 1.; nucleotides 95 to 244 in SEQ ID No.2, nucleotides 108 to 257 in SEQ ID No. 3 and nucleotides 104 to 253 in SEQ ID No. 4.

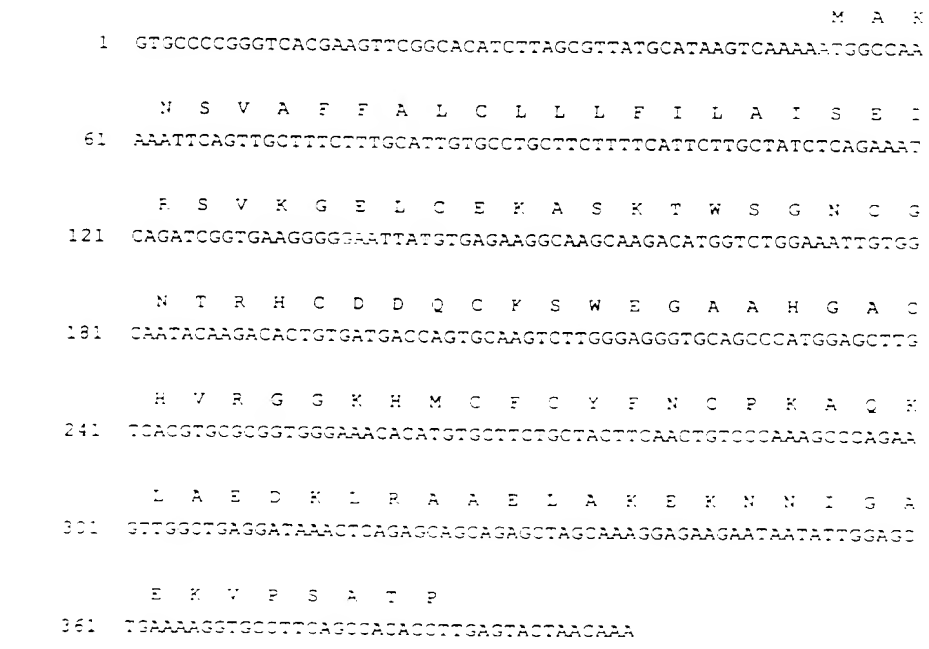
21. Use of the polynucleotide of any one of claims 1 to 10, or the vector of claim 11 in the production of an antimicrobial protein.

22. Use of any one of the polynucleotides of claim 4 or claim 5 as a signal sequence.

## Dm Gene Structure




## Dm2.18



Dm 2.1

```
                    M  A  K  N  S  V  A  F  L  A  F  L  L  L  L  F  V
    1  GGCACGAGTAATGGCCAAAAATTCAGTTGCTTTCTTAGCATTTCTTCTGCTTCTTTTCGT

                 L  A  I  S  E  I  G  S  V  K  G  E  L  C  E  K  A  S  K  T
   61  TCTTGCTATCTCAGAAATCGGATCGGTGAAGGGGGAATTATGTGAGAAGGCAAGCAAGAC

                 W  S  G  N  C  G  N  T  R  H  C  D  D  Q  C  K  S  W  E  G
  121  ATGGTCTGGAAATTGTGGCAATACAAGACACTGTGATGACCAGTGCAAGTCTTGGGAGGG

                 A  A  H  G  A  C  H  V  R  G  G  K  H  M  C  F  C  Y  F  N
  181  CGCAGCCCATGGAGCTTGTCACGTGCGCGGTGGGAAACACATGTGCTTTTGCTACTTCAA
       GCGTCGGGTACCTCGAACAGTGCACGCGCCACCCTTTGTGTACACGAAAACGATGAAGTT

                 C  S  K  A  Q  K  L  A  Q  D  K  L  K  A  D  K  L  A  K  E
  241  CTGTTCCAAAGCCCAGAAGCTGGCTCAGGATAAACTCAAAGCCGACAAGCTCGCCAAGGA

                 K  S  E  A  E  K  V  P  A  T  P
  301  GAAGAGTGAAGCCGAAAAGGTGCCAGCTACACCTTGAGTACTAACAAGTGTTGTATGATT

  361  ATGAATAAAGAGAAAATGCTTTCTAGTTACCATATTTAGCATTCTCTAATGTGTAATGTT

  421  TGTTGCTTTTGGAACTAATTGCTTAACTATGATTCCAGCTAATAATGTTTTAAGTATATA

  481  ATATAAGTTATCTTATTTTGAAGCCTGTAAAAAAAAAAAAAAA
```

Dm2.3

```
                          M  A  K  N  S  V  A  F  F  A  F  V
    1  CGGCACGAGGCACAATCTCAAAAATGGCCAAAAATTCGGTTGCTTTCTTTGCATTTGTCC

                 L  L  L  F  V  L  A  I  S  E  I  G  S  V  K  G  E  L  C  E
   61  TGCTTCTTTTCGTTCTTGCTATCTCAGAAATTGGATCGGTGAAGGGAGAATTATGTGAGA

                 K  A  S  K  T  W  S  G  N  C  G  I  T  S  H  C  D  N  Q  C
  121  AGGCAAGCAAGACATGGTCTGGAAATTGTGGCATCACATCACACTGTGACAACCAGTGCC

                 R  S  W  E  G  A  I  H  G  A  C  H  V  R  G  G  K  H  M  C
  181  GGTCGTGGGAGGGTGCAATCCATGGAGCTTGTCACGTGCGCGGTGGGAAACACATGTGCT

                 F  C  Y  F  N  C  S  K  A  D  E  L  A  K  E  K  I  E  A  E
  241  TGTGCTACTTCAACTGTTCCAAAGCCGATGAGCTCGCGAAGGAGAAGATTGAAGCCGAAA

                 K  M  P  A  T  P
  301  AGATGCCAGCCACACCTTGAGTACTAACAAATGCTATATGATTATAAATAAAGAGAAAAT
```

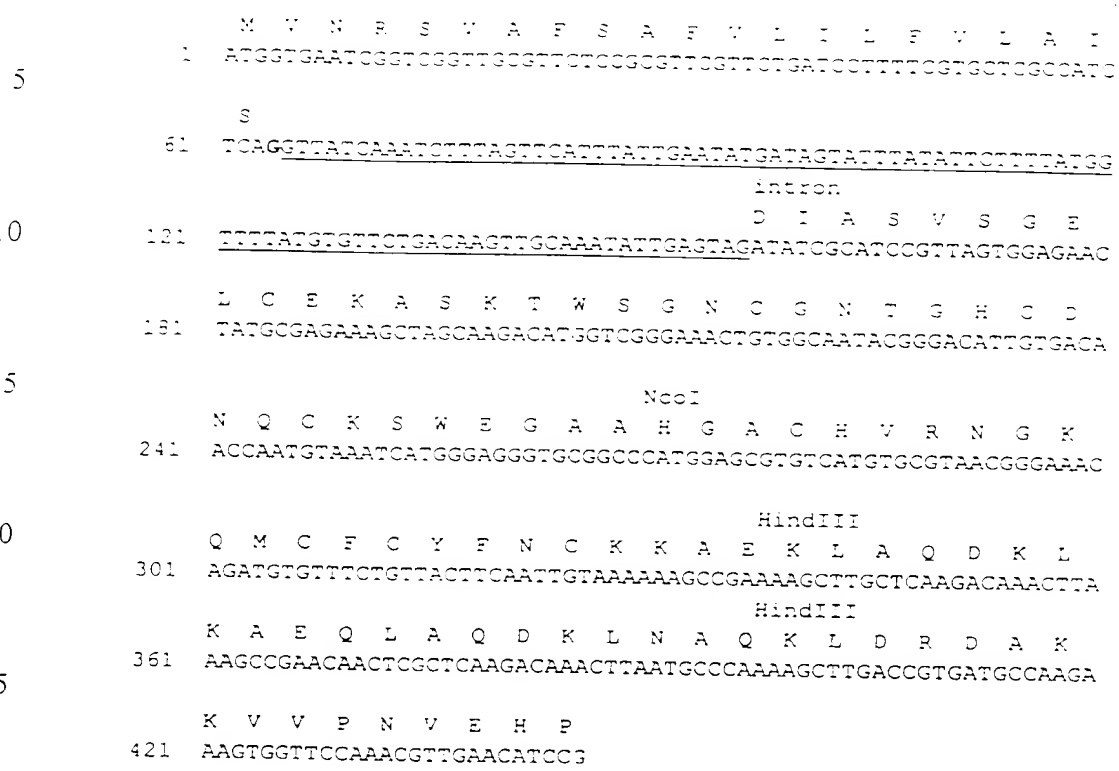
361 GCTTTGTAAAAAAAAAAAAAAAAA

Dm2.5

```
                                    M  V  N  R  S  V  A  F  S  V  F  V  L  I
    1 GGCACGAGCCTATTAAAAAATGGTGAATCGATCGGTTGCTTTCTCCGTGTTCGTTCTGAT

                 L  F  V  L  A  I  S  D  I  T  S  V  R  G  E  V  C  E  K  A
   61 CCTTTTCGTGCTCGCCATCTCAGATATCACAAGTGTGAGAGGAGAAGTATGCGAGAAAGC

                 S  K  T  W  S  G  N  D  G  N  T  G  H  C  D  N  Q  C  K  Y
  121 TAGCAAGACATGGTCAGGAAACTGTGGCAACACGGGACACTGTGACAACCAATGTAAATA

                 W  E  G  A  A  H  G  A  C  H  V  R  G  G  K  H  M  C  F  C
  181 CTGGGAGGGGGCGGCCCATGGGGCGTGCCACGTGCGTGGAGGGAAACACATGTGTTTCTG

                 Y  F  K  C  P  K  A  E  K  L  A  Q  D  K  V  N  A  Q  E  L
  241 CTACTTCAAGTGTCCCAAAGCCGAAAAGCTTGCTCAAGACAAAGTTAATGCCCAAGAGCT

                 D  R  D  A  K  K  V  I  P  N  V  E  H  P
  301 TGACCGTGATGCCAAGAAAGTGATTCCGAACGTTGAACATCCGTGAAAGGGTCGGTTTCT

  361 TTAAATAGAAAGTCTTAGATTACGAATGCGAATAACTATAGAAAATGTTTGCTAAATGTC

  421 ACATTATAATTAGAACTTTATGATTGTTGTCAATAGGGCATTTTCTTGTTAGTGATATGT

  481 GTAATAAGGTGATGCTTTTATGCTTTTCGTGCGTAAGAGTTTTCGACTATGTGTAATAAA

  541 GAAAGGGTCTTTTTTTTTTAAAAAAAAAAAAAAAAAA
```

Figure 1

A) Dm Gene Structure Sequence ID No. 5




B) Dm2.18 Sequence ID No. 1

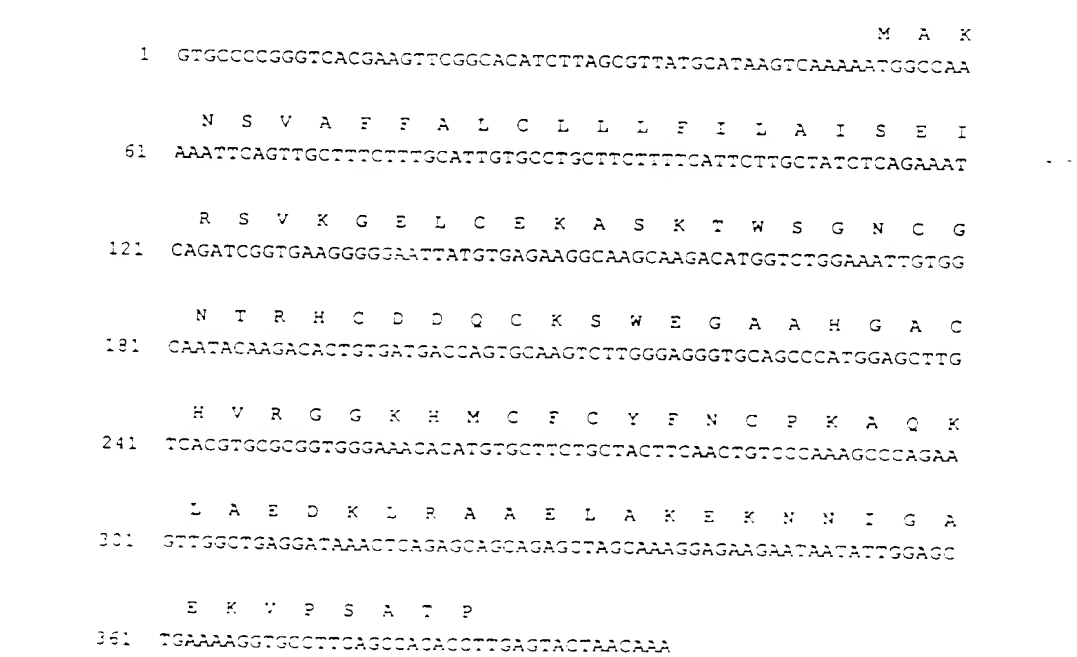


Figure 2

Dm2.1 Sequence ID No. 2

```
                    M  A  K  N  S  V  A  F  L  A  F  L  L  L  L  F  V
    1  GGCACGAGTAATGGCCAAAAATTCAGTTGCTTTCTTAGCATTTCTTCTGCTTCTTTTCGT

                 L  A  I  S  E  I  G  S  V  K  G  E  L  C  E  K  A  S  K  T
   61  TGTTGCTATCTCAGAAATCGGATCGGTGAAGGGGGAATTATGTGAGAAGGCAAGCAAGAC

                 W  S  G  N  C  G  N  T  R  H  C  D  D  Q  C  K  S  W  E  G
  121  ATGGTCTGGAAATTGTGGCAATACAAGACACTGTGATGACCAGTGCAAGTCTTGGGAGGG

                 A  A  H  G  A  C  H  V  R  G  G  K  H  M  C  F  C  Y  F  N
  181  CGCAGCCCATGGAGCTTGTCACGTGCGCGGTGGGAAACACATGTGCTTTTGCTACTTCAA
       GCGTCGGGTACCTCGAACAGTGCACGCGCCACCCTTTGTGTACACGAAAACGATGAAGTT

                 C  S  K  A  Q  K  L  A  Q  D  K  L  K  A  D  K  L  A  K  E
  241  CTGTTCCAAAGCCCAGAAGCTGGCTCAGGATAAACTCAAAGCCGACAAGCTCGCCAAGGA

                 K  S  E  A  E  K  V  P  A  T  P
  301  GAAGAGTGAAGCCGAAAAGGTGCCAGCTACACCTTGAGTACTAACAAGTGTTGTATGATT

  361  ATGAATAAAGAGAAAATGCTTTCTAGTTACCATATTTAGCATTCTCTAATGTGTAATGTT

  421  TGTTGCTTTTGGAACTAATTGCTTAACTATGATTCCAGCTAATAATGTTTTAAGTATATA

  481  ATATAAGTTATCTTATTTTGAAGCCTGTAAAAAAAAAAAAAAA
```

Dm2.3 Sequence ID No. 3

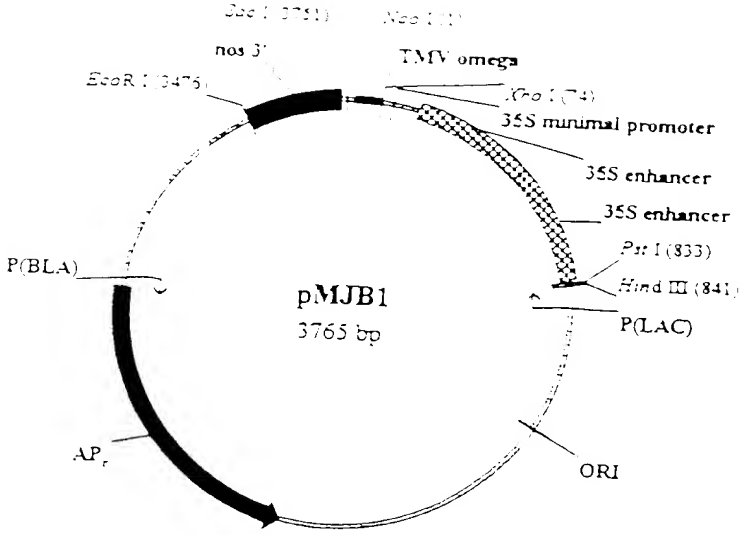
```
                          M  A  K  N  S  V  A  F  F  A  F  V
    1  CGGCACGAGGCACAATCTCAAAAATGGCCAAAAATTCGGTTGCTTTCTTTGCATTTGTCC

                 L  L  L  F  V  L  A  I  S  E  I  G  S  V  K  G  E  L  C  E
   61  TGCTTCTTTTCGTTCTTGCTATCTCAGAAATTGGATCGGTGAAGGGAGAATTATGTGAGA

                 K  A  S  K  T  W  S  G  N  C  G  I  T  S  H  C  D  N  Q  C
  121  AGGCAAGCAAGACATGGTCTGGAAATTGTGGCATCACATCACACTGTGACAACCAGTGCC

                 R  S  W  E  G  A  I  H  G  A  C  H  V  R  G  G  K  H  M  C
  181  GGTCGTGGGAGGGTGCAATCCATGGAGCTTGTCACGTGCGCGGTGGGAAACACATGTGCT

                 F  C  Y  F  N  C  S  K  A  D  E  L  A  K  E  K  I  E  A  E
  241  TCTGCTACTTCAACTGTTCCAAAGCCGATGAGCTCGCGAAGGAGAAGATTGAAGCCGAAA

                 K  M  P  A  T  P
  301  AGATGCCAGCCACACCTTGAGTACTAACAAATGCTATATGATTATAAATAAAGAGAAAAT

  361  GCTTTCTAAAAAAAAAAAAAAAAAAAA
```

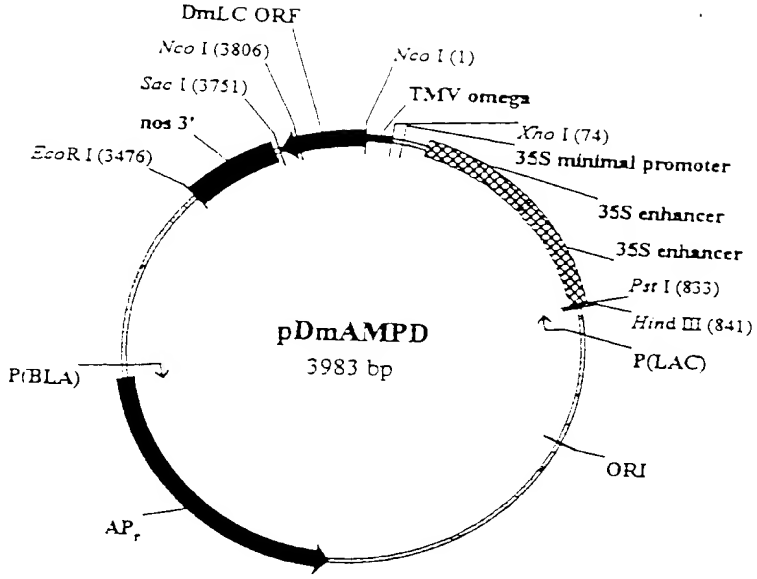
Figure 3

Dm2.5 Sequence ID No. 4

```
                                                M  V  N  R  S  V  A  F  S  V  F  V  L  I
   1  GGCACGAGCCTATTAAAAAAATGGTGAATCGATCGGTTGCTTTCTCCGTGTTCGTTCTTGAT

                L  F  V  L  A  I  S  D  I  T  S  V  R  G  E  V  C  E  K  A
  61  CCTTTTCGTGCTCGCCATCTCAGATATCACAAGTGTGAGAGGAGAAGTATGCGAGAAAGC

                S  K  T  W  S  G  N  C  G  N  T  G  H  C  D  N  Q  C  K  Y
 121  TAGCAAGACATGGTCAGGAAACTGTGGCAACACGGGACACTGTGACAACCAATGTAAATA

                W  E  G  A  A  H  G  A  C  H  V  R  G  G  K  H  M  C  F  C
 181  CTGGGAGGGGGCGGCCCATGGGGCGTGCCACGTGCGTGGAGGGAAACACATGTGTTTCTG

                Y  F  K  C  P  K  A  E  K  L  A  Q  D  K  V  N  A  Q  E  L
 241  CTACTTCAAGTGTCCCAAAGCCGAAAAGCTTGCTCAAGACAAAGTTAATGCCCAAGAGCT

                D  R  D  A  K  K  V  I  P  N  V  E  H  P
 301  TGACCGTGATGCCAAGAAAGTGATTCCGAACGTTGAACATCCGTGAAAGGGTCGGTTTCT

 361  TTAAATAGAAAGTCTTAGATTACGAATGCGAATAACTATAGAAAATGTTTGCTAAATGTC

 421  ACATTATAATTAGAACTTTATGATTGTTGTCAATAGGGCATTTTCTTGTTAGTGATATGT

 481  GTAATAAGGTGATGCTTTTATGCTTTTCGTGCGTAAGAGTTTTCGACTATGTGTAATAAA

 541  GAAAGGGTCTTTTTTTTTTTAAAAAAAAAAAAAAAAAA
```

Figure 4


Sac I (3751)
Nco I (1)
nos 3'
TMV omega
EcoR I (3476)
Xho I (74)
35S minimal promoter
35S enhancer
35S enhancer
Pst I (833)
Hind III (841)
P(BLA)
pMJB1
3765 bp
P(LAC)
AP$_r$
ORI


DmLC ORF
Nco I (3806)
Nco I (1)
Sac I (3751)
TMV omega
nos 3'
Xho I (74)
EcoR I (3476)
35S minimal promoter
35S enhancer
35S enhancer
Pst I (833)
Hind III (841)
pDmAMPD
3983 bp
P(BLA)
P(LAC)
ORI
AP$_r$

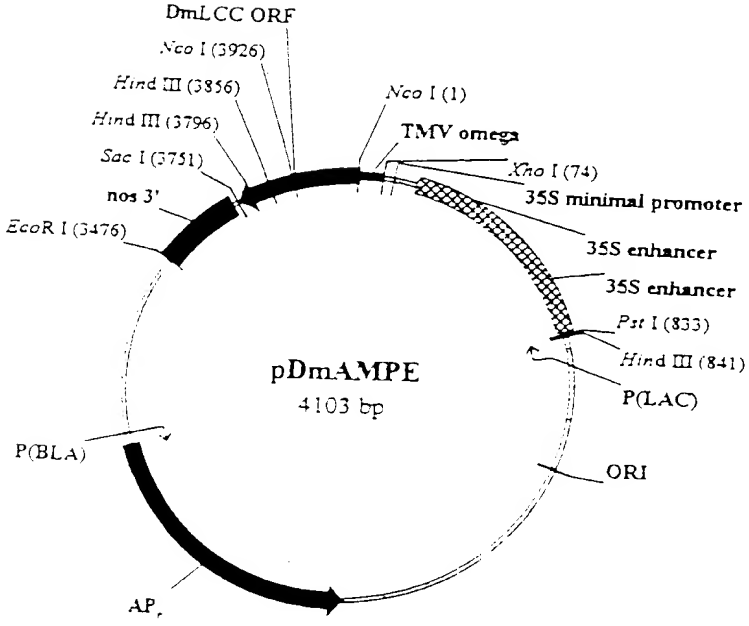
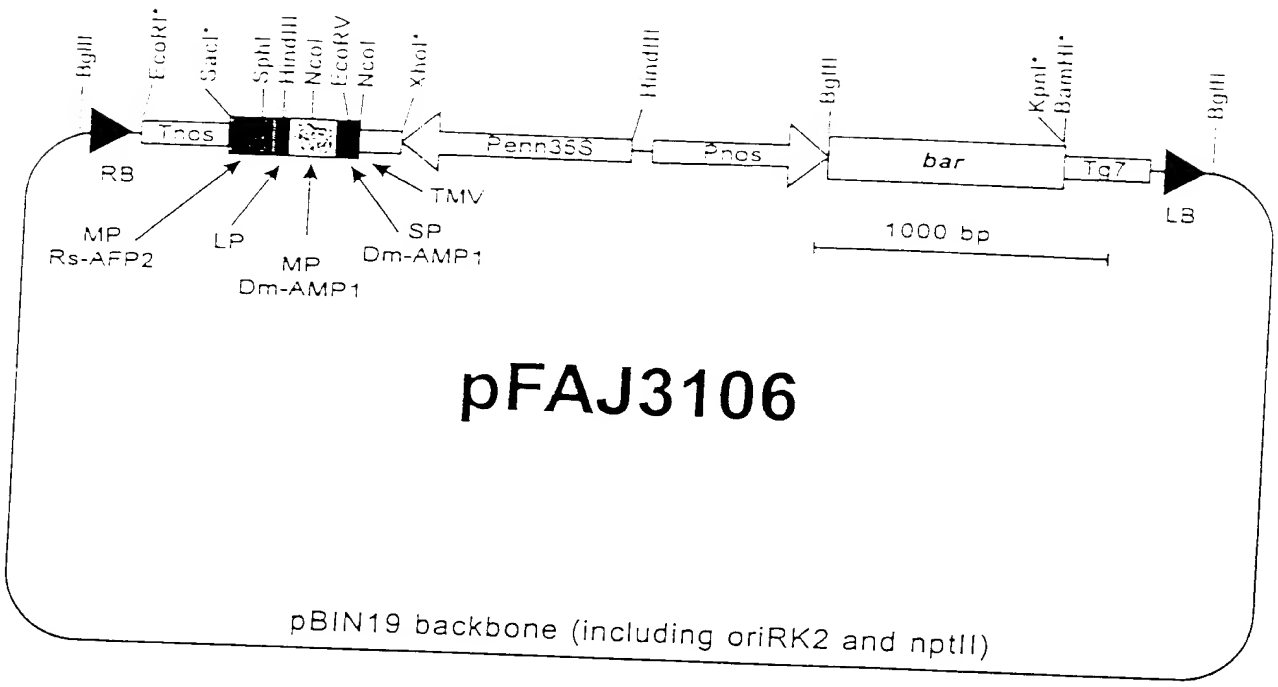

DmLCC ORF
Nco I (3926)
Hind III (3856)
Nco I (1)
Hind III (3796)
TMV omega
Sac I (3751)
Xho I (74)
nos 3'
35S minimal promoter
EcoR I (3476)
35S enhancer
35S enhancer
Pst I (833)
Hind III (841)
pDmAMPE
4103 bp
P(LAC)
P(BLA)
ORI
AP$_r$

Figure 5

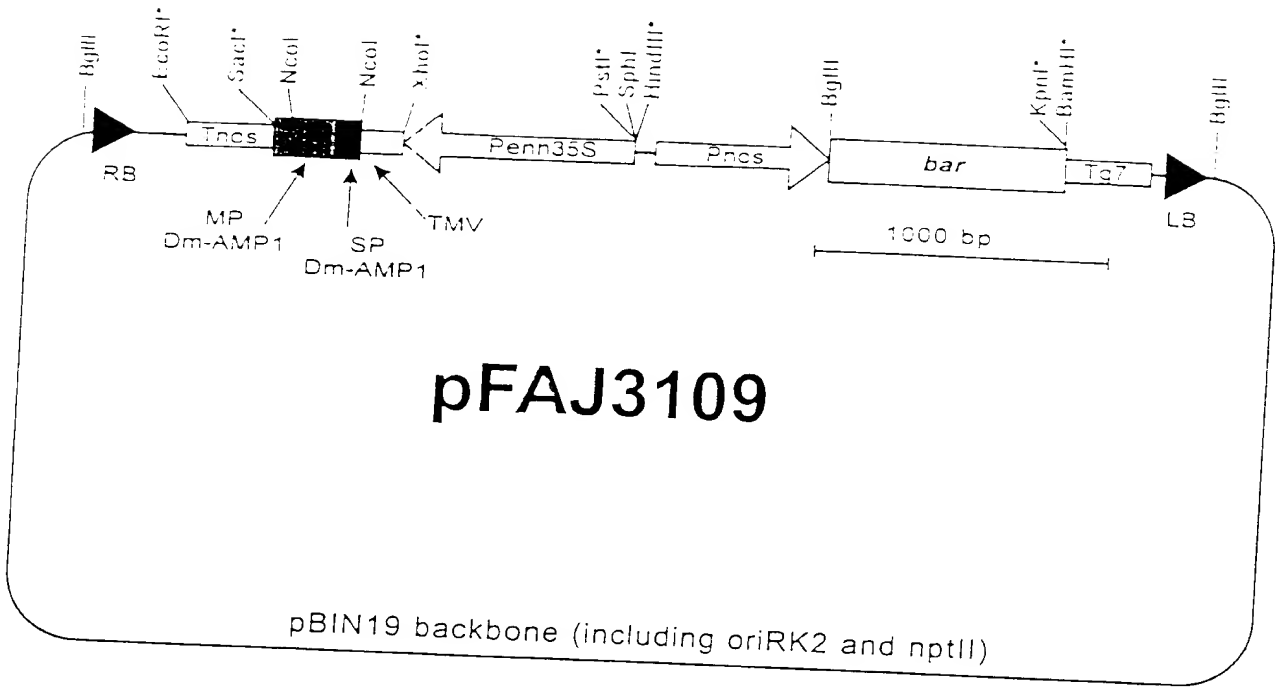



## Symbols

RB: right border of T-DNA
Tnos: terminator of T-DNA nopaline synthase gene
MP Rs-AFP2: mature protein domain of Rs-AFP2
LP: first 16 AA of Dm-AMP1 C-terminal propeptide and subtilisin-like protease recognition site IGKR
MP Dm-AMP1: mature protein domain of Dm-AMP1 cDNA
SP Dm-AMP1: signal peptide domain of Dm-AMP1 cDNA
TMV: tobacco mosaic virus 5' leader sequence
Penh35S: promotor of 35S RNA of cauliflower mosaic virus with duplicated enhancer region
Pnos: promotor of T-DNA nopaline synthase gene
bar: basta resistance encoding gene
Tg7: terminator of T-DNA gene 7
LB: left border of T-DNA

*: unique restriction site

Figure 6




## Symbols

RB: right border of T-DNA
Tnos: terminator of T-DNA nopaline synthase gene
MP Dm-AMP1: mature protein domain of Dm-AMP1
SP Dm-AMP1: signal peptide domain of Dm-AMP1 cDNA
TMV: tobacco mosaic virus 5' leader sequence
Penh35S: promotor of 35S RNA of cauliflower mosaic virus with duplicated enhancer region
Pnos: promotor of T-DNA nopaline synthase gene
bar: basta resistance encoding gene
Tg7: terminator of T-DNA gene 7
LB: left border of T-DNA

*: unique restriction site

Figure 7

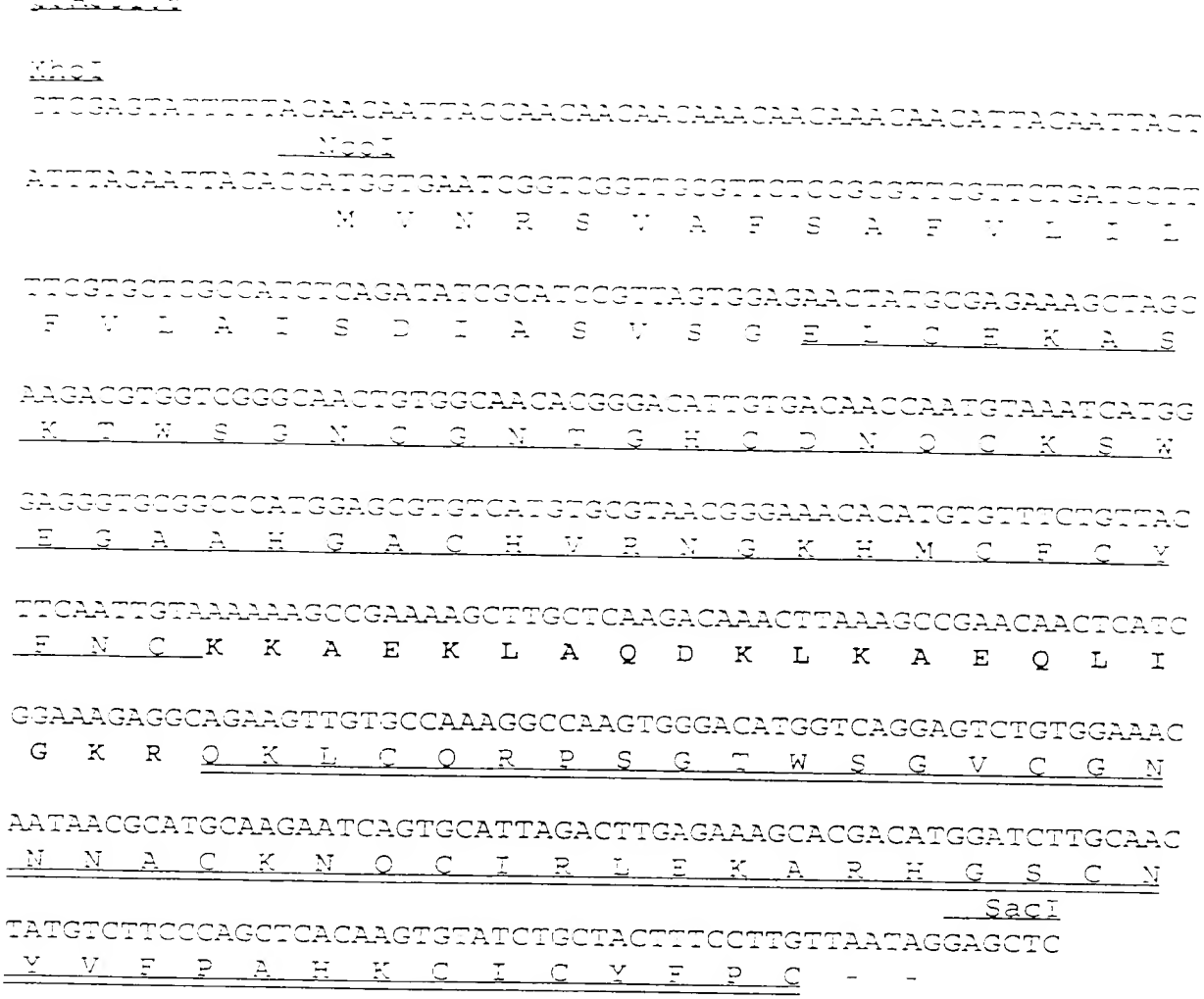


Figure 8

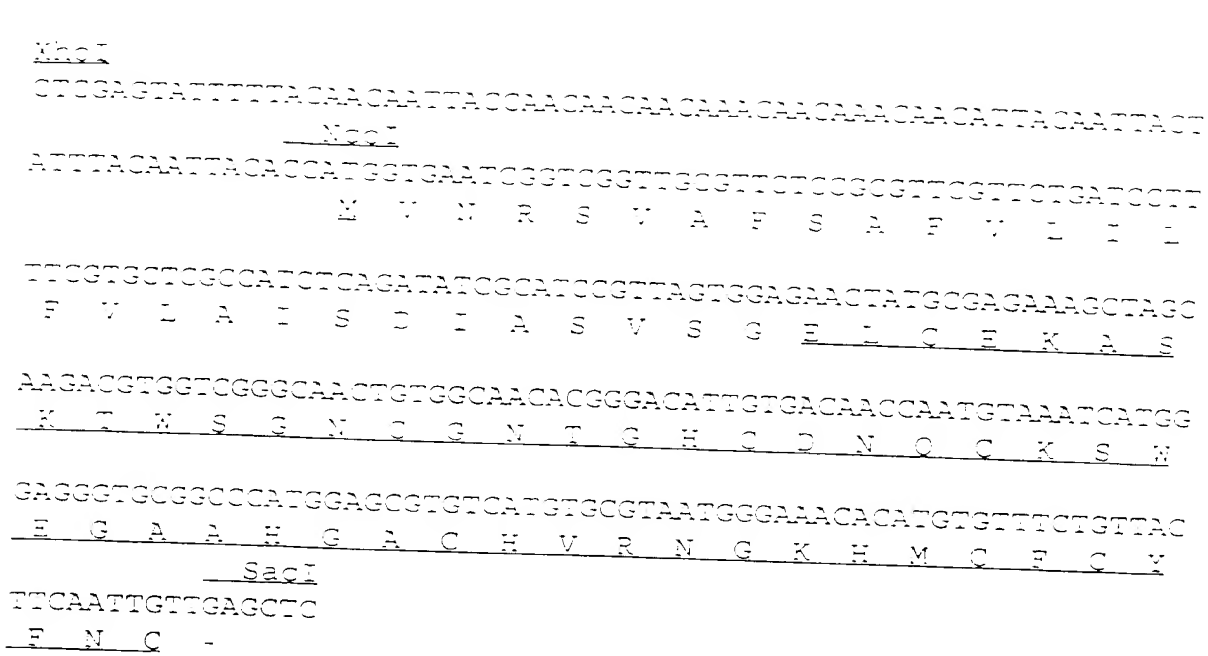

XhoI
NcoI
M V N R S V A F S A F V L I L
F V L A I S D I A S V S G E L C E K A S
K T W S G N C G N T G H C D N Q C K S W
E G A A H G A C H V R N G K H M C F C V
SacI
F N C -

Figure 9

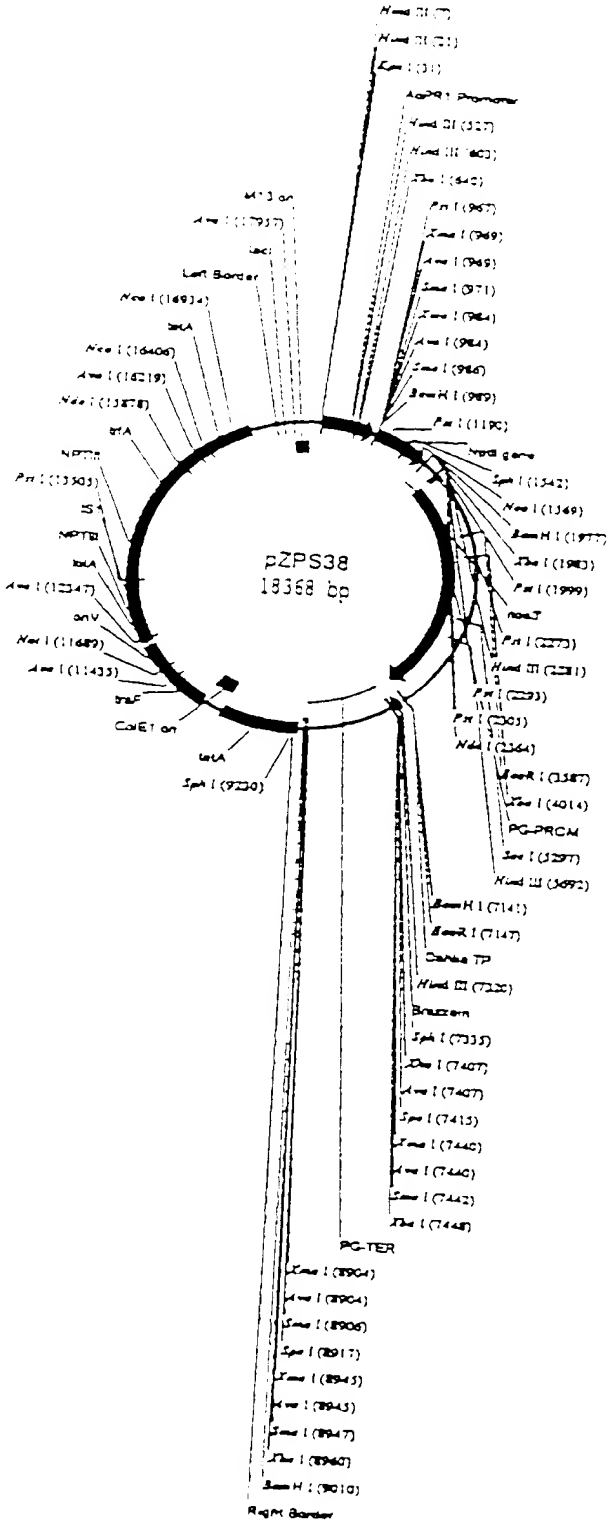

pZPS38
18368 bp
Left Border
NPTII
ori V
ColE1 ori
PG-PRCM
PG-TER
Right Border

Figure 10

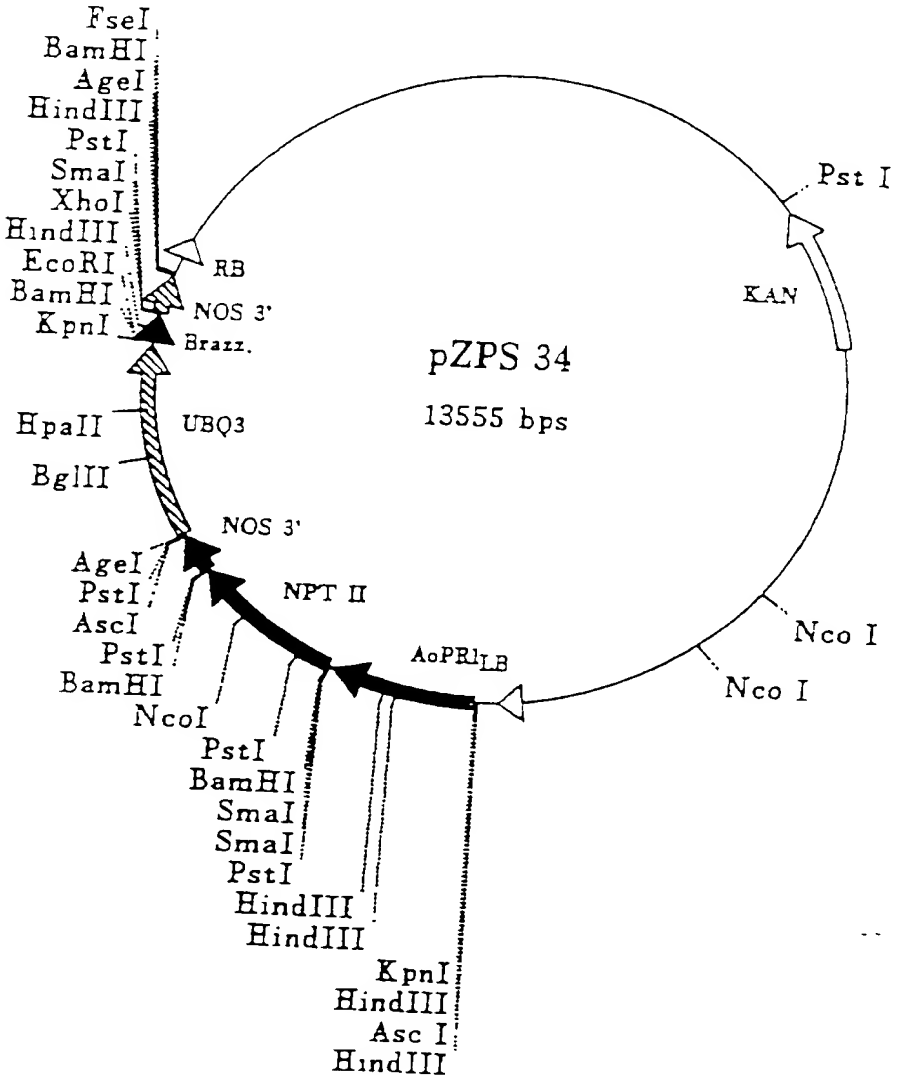

FseI
BamHI
AgeI
HindIII
PstI
SmaI
XhoI
HindIII
EcoRI
BamHI
KpnI
RB
NOS 3'
Brazz.
HpaII
UBQ3
BglII
NOS 3'
AgeI
PstI
AscI
PstI
BamHI
NcoI
PstI
BamHI
SmaI
SmaI
PstI
HindIII
HindIII
KpnI
HindIII
Asc I
HindIII
NPT II
AoPRI LB
pZPS 34
13555 bps
Pst I
KAN
Nco I
Nco I

Figure 11

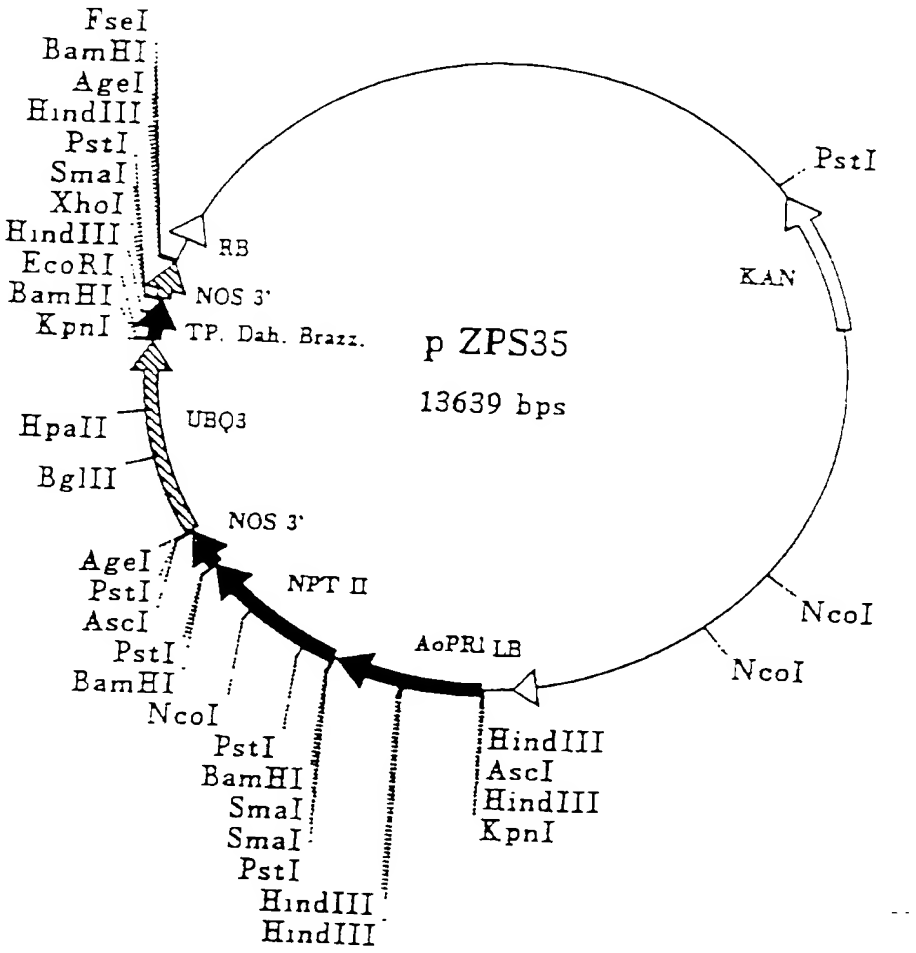
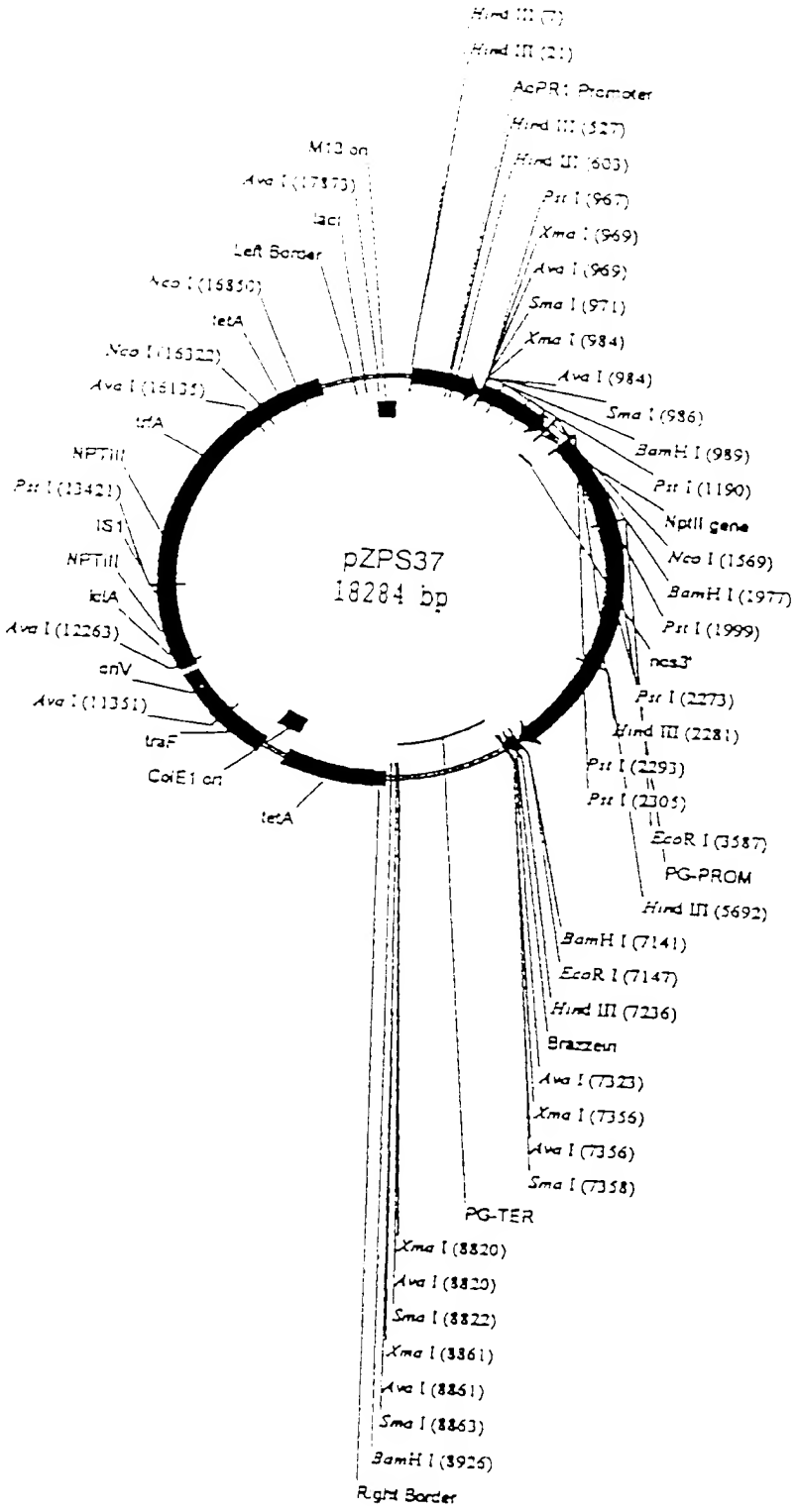
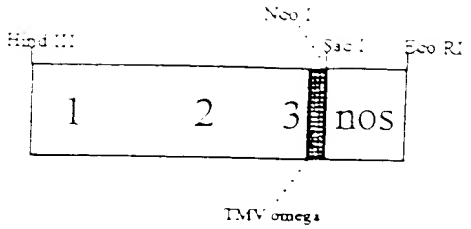

FseI
BamHI
AgeI
HindIII
PstI
SmaI
XhoI
HindIII
EcoRI
BamHI
KpnI
RB
NOS 3'
TP. Dah. Brazz.
p ZPS35
13639 bps
PstI
KAN
HpaII
BglII
UBQ3
NOS 3'
AgeI
PstI
AscI
PstI
BamHI
NcoI
NPT II
AoPR1
LB
NcoI
NcoI
PstI
BamHI
SmaI
SmaI
PstI
HindIII
HindIII
HindIII
AscI
HindIII
KpnI

Figure 12


pZPS37
18284 bp
Hind III (7)
Hind III (21)
AcPR1 Promoter
Hind III (527)
Hind III (603)
Pst I (967)
Xma I (969)
Ava I (969)
Sma I (971)
Xma I (984)
Ava I (984)
Sma I (986)
BamH I (989)
Pst I (1190)
NptII gene
Nco I (1569)
BamH I (1977)
Pst I (1999)
nos3'
Pst I (2273)
Hind III (2281)
Pst I (2293)
Pst I (2305)
EcoR I (3587)
PG-PROM
Hind III (5692)
BamH I (7141)
EcoR I (7147)
Hind III (7236)
Brazzein
Ava I (7323)
Xma I (7356)
Ava I (7356)
Sma I (7358)
PG-TER
Xma I (8820)
Ava I (8820)
Sma I (8822)
Xma I (8861)
Ava I (8861)
Sma I (8863)
BamH I (8926)
Right Border
tetA
ColE1 ori
traF
Ava I (11351)
oriV
Ava I (12263)
tetA
IS1
NPTIII
Pst I (13421)
NPTIII
tetA
Ava I (15135)
Nco I (15322)
tetA
Nco I (15850)
Left Border
lacI
Ava I (17873)
M13 ori

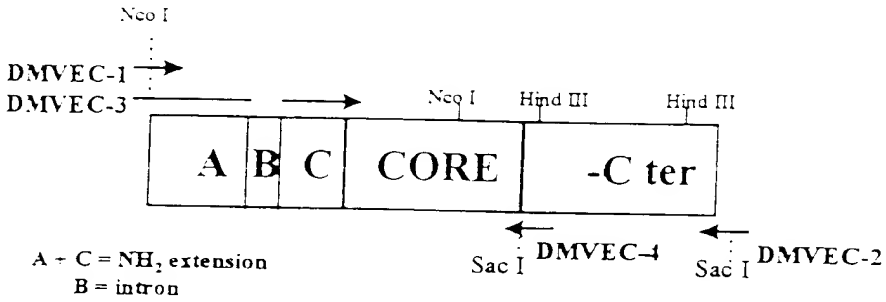
p[illegible]JB1

Figure 13



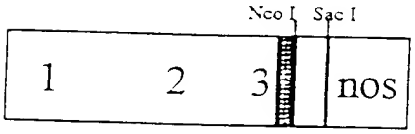

1,2 = CaMV35S enhancer
3 = CaMV35S minimal promoter

## Structure of DmAMP1 Gene and position of vector construction oligonucleotides

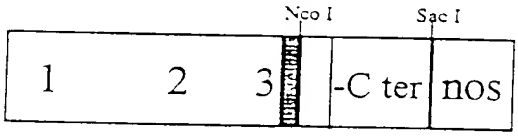



PCR genomic DNA with DMVEC 1 and 4. Isolate 60 bp Nco I/Sac I fragment, clone into pMJB1 Nco I/Sac I = pDmAMPA

PCR genomic DNA with DMVEC 1 and 2. Isolate 176 bp Nco I/Sac I fragment, clone into pMJB1 Nco I/Sac I = pDmAMPB

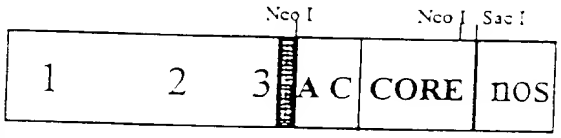



**pDmAMPA**

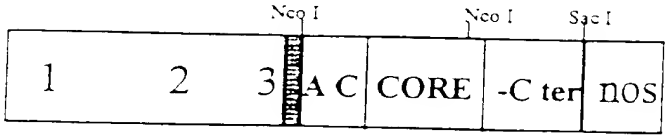



**pDmAMPB**

PCR genomic DNA with DMVEC 3 and 4. Isolate 150 bp Nco I fragment, clone into pDmAMPA and pDmAMPB Nco I = pDmAMPD and pDmAMPE




**pDmAMPD**




**pDmAMPE**

Figure 14

Sequence ID No. 6 Dm-AMP1

GAG CTT TGC GAG AAG GCT TCT AAG ACT TGG TCT GGA AAC

TGG GAG GGA GCT GCT CAT GGA GCT TGC CAT GTT AGA AAC

Sequence ID No. 7 Dm-AMP2

GAG GTT TGC GAG AAG GCT TCT AAG ACT TGG TCT GGA AAC

Figure 14 Continued

TGC GGA AAC ACT GGA CAT TGC GAT AAC CAA TGC AAG TCT

GGA AAG CAT ATG TGC TTC TGC TAC TTC AAC TGC

TGC GGA AAC ACT GGA CAT TGC ... ... ...